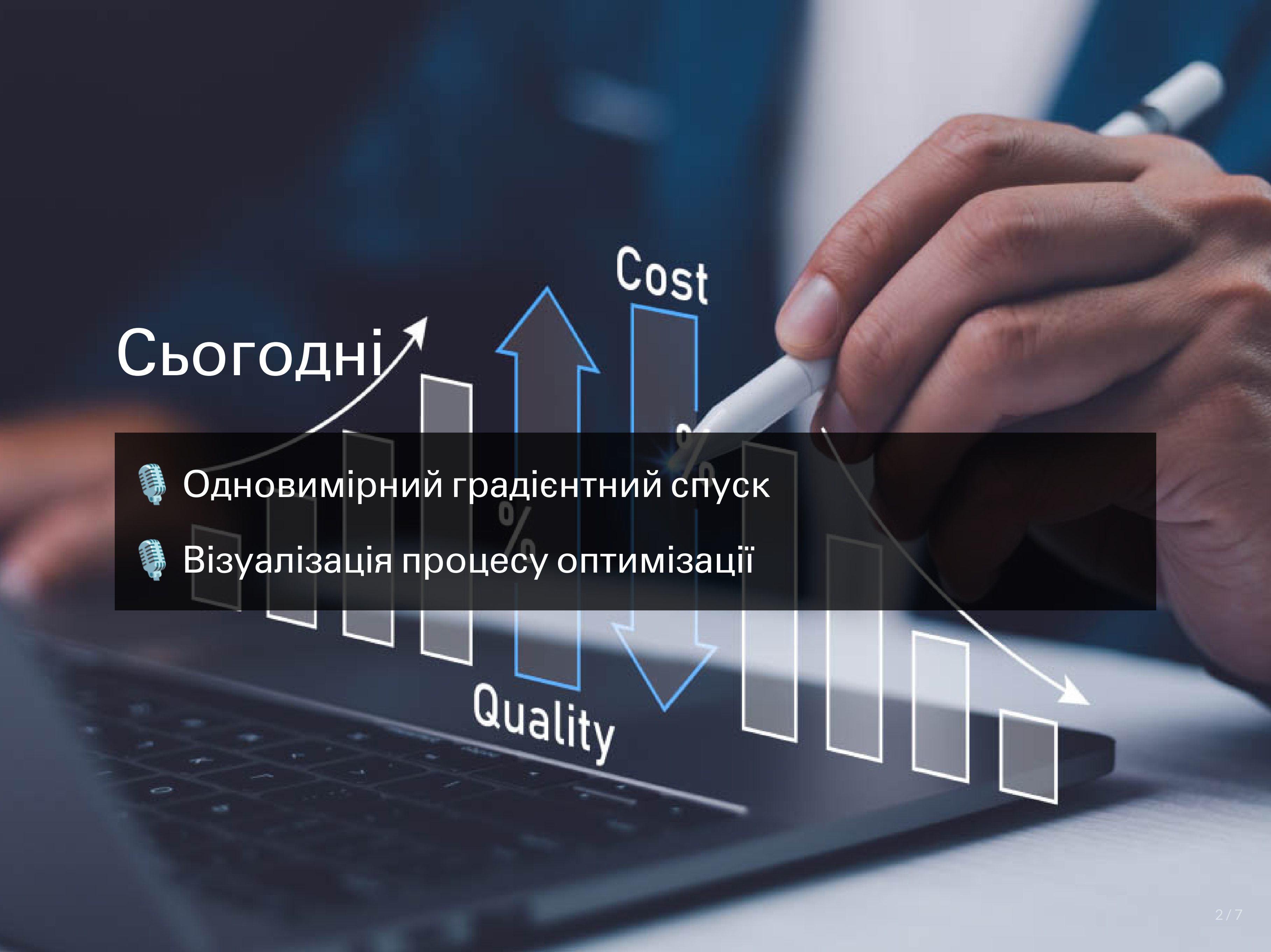
Методи чисельної оптимізації

Лекція 4: Одновимірний градієнтний спуск

Кочура Юрій Петрович
iuriy.kochura@gmail.com
@y_kochura



Сьогодні

- 🎙️ Одновимірний градієнтний спуск
- 🎙️ Візуалізація процесу оптимізації

Одновимірний градієнтний спуск

Одновимірний градієнтний спуск

Розглянемо деяку монотонну неперервну диференційовану функцію $f : \mathbb{R} \to \mathbb{R}$. Розкладаючи у ряд Тейлора, ми отримуємо:

$$f(x+\varepsilon) = f(x) + \varepsilon f^{'}(x) + \mathcal{O}(\varepsilon^2)$$

Зфіксуємо розмір кроку $\alpha > 0$ та оберемо $\varepsilon = -\alpha f^{'}(x)$. Підставляючи це у ряд Тейлора, отримаємо:

$$f(x - \alpha f^{'}(x)) = f(x) - \alpha f^{'2}(x) + \mathcal{O}(\alpha^2 f^{'2}(x))$$

Якщо похідна $f^{'}(x) \neq 0$ не зникає ми робимо прогрес так як $\alpha f^{'2}(x) > 0$. Крім того, ми завжди можемо вибрати α досить малим, щоб вирази вищих порядків стали нерелевантними. Тому ми приходимо до

$$f(x - \alpha f^{'}(x)) \lessapprox f(x)$$

Це означає, якщо ми використовуємо

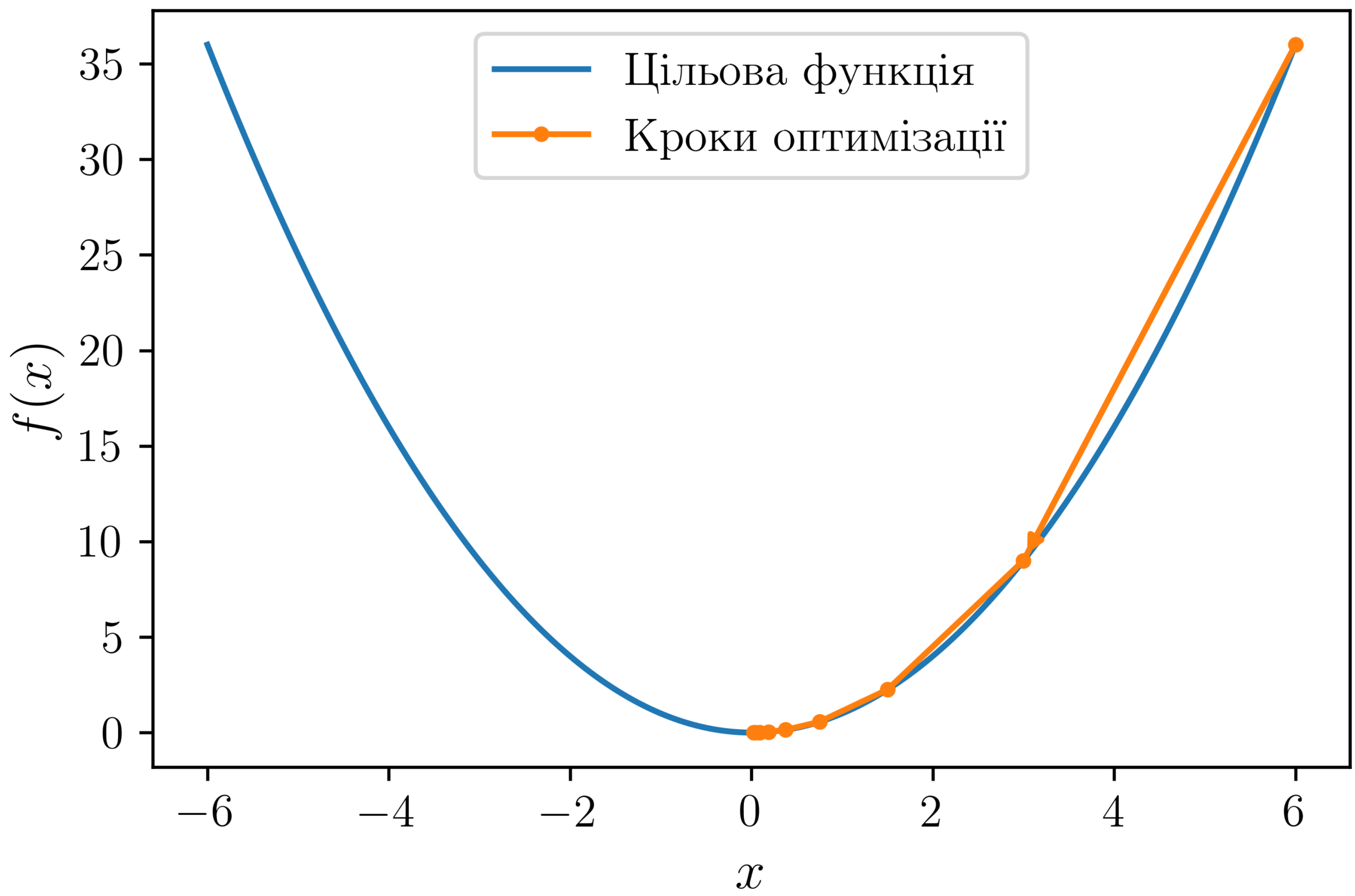
$$x \leftarrow x - \alpha f^{'}(x)$$

для ітерації по x, значення функції $f(x)$ може зменшитись.

Візуалізація процесу оптимізації

```python
import numpy as np

def f(x):  # Objective function
    return x**2

def f_grad(x):  # Gradient (derivative) of the objective function
    return 2 * x

def bgd(alpha, f_grad):
    x = 6.0           # Initial value of x
    results = [x]
    epoch = 8         # Number of iterations
    for i in range(epoch):
        x -= alpha * f_grad(x)
        results.append(float("%.6f" % x))
    print(f'epoch {epoch}, x: {x:.6f}')
    return results

results = bgd(0.25, f_grad)
print(results)
```

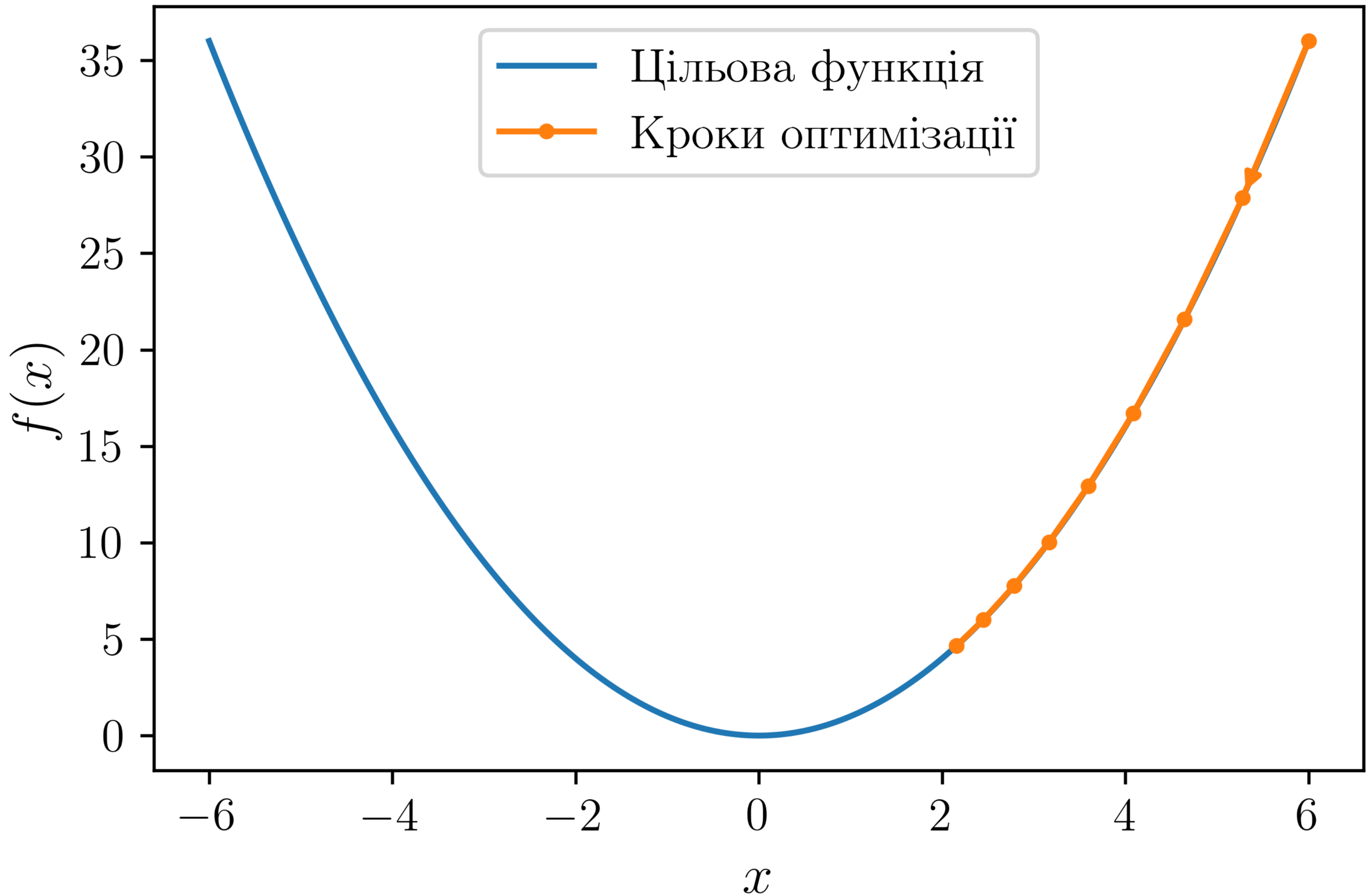
```
epoch 8, x: 0.023438
[6.0, 3.0, 1.5, 0.75, 0.375, 0.1875, 0.09375, 0.046875, 0.023438]
```



Одновимірний градієнтний спуск ($\alpha = 0.25$)

```
results = bgd(0.06, f_grad)
```

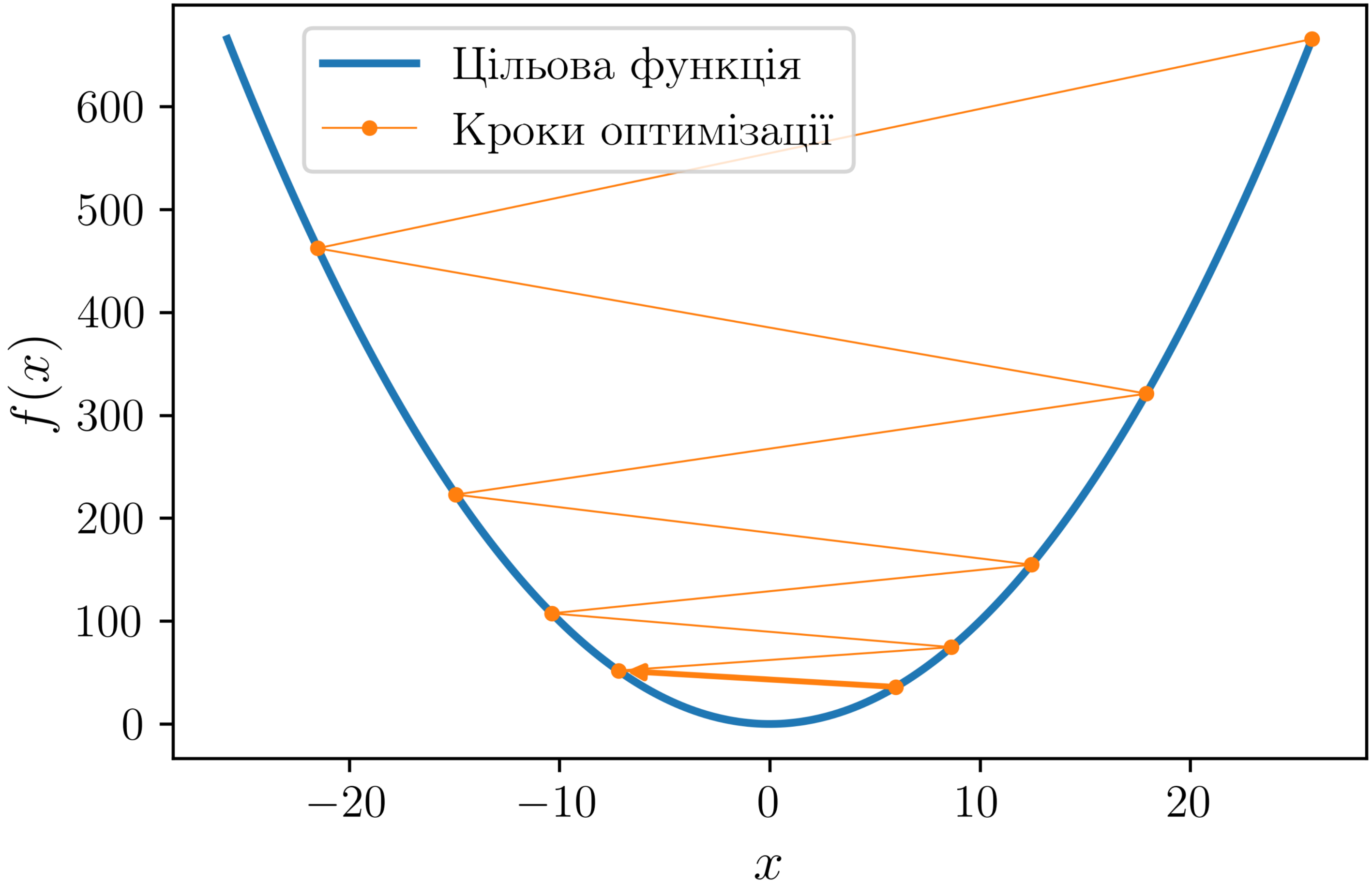
```
epoch 8, x: 2.157807
```



Одновимірний градієнтний спуск ($\alpha = 0.06$)

```
results = bgd(1.1, f_grad)
```

```
epoch 8, x: 25.798902
```



Одновимірний градієнтний спуск ($\alpha = 1.1$)